

Fourier-transzformációs közép-infravörös spektroszkópia alapú szervesanyag-tartalom becslés tábla szintű reprezentativitás-vizsgálata kemometriai módszerekkel

*TÓTH József Attila, DÖBRÖNTEY Réka, SZEGI Tamás, MICHÉLI Erika, CSORBA Ádám

Magyar Agrár- és Élettudományi Egyetem, Gödöllő

(Beérkezett: 2020.09.21.; Elfogadva: 2021.04.22.)

Bevezetés

A talaj egy rendkívül komplex fizikai, kémiai és biológiai és ökológiai folyamatokkal jellemezhető rendszer, összefüggéseinek részletes ismerete több okból is kiemelten fontos (TAN, 1982). Amellett, hogy bolygónk szabadföldi mezőgazdaságának, így az élelmiszerláncnak alapja, a globális szénkörü forgásban is meghatározó szerepet tölt be (RAMESH et al., 2019; LAL, 2013). Az egyes talajparaméterek valós idejű, pontos felmérése elengedhetetlen a bennük zajló folyamatok megfelelő értelmezése érdekében. Ezt célozzák a rendelkezésre álló talajtérképek, valamint a talajtérképezési módszertan fejlesztésével foglalkozó számos hazai tudományos munka is (DOBOS et al., 2010, 2011, 2014, 2019; PÁSZTOR et al., 2010, 2013, 2014, 2015, 2019). A precíziós gazdálkodás, valamint a környezetmodellezés jövőbeni térnyerésében kulcsszerepe van a talajokról rendelkezésre álló információ mennyiségének és minőségének.

A talaj funkcióinak fenntarthatósága szempontjából meghatározó szerepet tölt be annak szerves anyaga (WEIL & BRADY, 2017). A talaj aktív kibocsátóként és elnyelőként is jelentős eleme a globális szénkörü forgásnak, e mellett a talajművelésen keresztül beavatkozási lehetőségeink is nyílnak (CHENU et al., 2019). A talajoknak a szénkörü forgásban betöltött meghatározó szerepe miatt (LAL, 2010, 2017; LAL et al., 2015) a szervesszén talajműveléssel történő megkötését számos nemzetközi program (CIRCASA, 4p1000, CCI projektek, FCCT) is szorgalmazza. Ezért a talaj szervesszén viszonyainak megismerése, annak pontos feltérképezése elengedhetetlen egy szénmegőrzési stratégia kialakításához, ellenőrzéséhez. Az ehhez szükséges, nagyobb léptékű (0,2–2 km²) talajtérképezést megelőzően elsőként a térképezési módszertan fejlesztése és kisebb léptékben (0,01–0,2 km²) történő tesztelése szükséges (PÁSZTOR et al., 2017; SZATMÁRI & PÁSZTOR, 2019; SZATMÁRI et al., 2019).

Minden szervesanyag-tartalom térkép alapját továbbra is az adott területre vonatkozóan reprezentatív pontokról származó és megfelelő analitikai módszerrel vizsgált talajminták alkotják. A talajmintavételezés hagyományos módszerei közvetlen, megbízható adatokat szolgáltatnak, azonban idő- és költségigényesek, nagy területet érintő talajtérképezés során jelentős anyagi forrásokat igényelhetnek (SZABÓ et al., 2005). Bár a hagyományos módszertan a jövőben is várhatóan alapját

*Levelező szerző: TÓTH JÓZSEF ATTILA, Magyar Agrár- és Élettudományi Egyetem, 2100, Gödöllő, Páter Károly u. 1.
E-mail: toth.jozsef.attila@uni-mate.hu

fogja képezni a talajviszonyok felmérésének, a térképezés módszertana jelentős fejlődésen ment keresztül az utóbbi két évtizedben (DOBOS et al., 2007; SZATMÁRI & PÁSZTOR, 2016, 2019). A nagyobb léptékű talajtérképezésben jelentős előrelépést eredményezett a távérzékelési módszerek fejlődése (VISCARRA ROSSEL & ADAMCHUK, 2013; DEERY et al., 2014). Az alkalmazott geostatistikai modellek hozzájárulnak a mintavételi stratégia optimalizálásához, ezáltal csökkentik az adott területről szükséges talajminták számát, illetve maximalizálják a meglévő talajmintákból levonható következtetések mértékét. (PAHLAVAN RAD et al., 2014; SZATMÁRI et al., 2015). A műholdak által és a drónokkal készített felvételekből származó hatalmas mennyiségű képi adat felhasználásának köszönhetően, a modern módszerekkel készült térképek jóval nagyobb számban lettek elérhetők. Ezen technológiák hatására a nagy méretarányú (regionális léptékű) talajtérképek készítéséhez és fejlesztéséhez felhasznált anyagi erőforrások mindeközben jelentősen csökkenhettek (BALLABIO et al., 2016). Mindemellett kisebb léptékben, nagyobb adatsűrűség mellett eltérő szempontok szerint történik az adatok értékelése, így a precíziós gazdálkodás alapját jelentő tábla szintű talajtérképek készítése is. Az elérhető nagyobb adatsűrűségnek köszönhetően jobban vizsgálhatók az összefüggések a hagyományos, illetve a közvetett vizsgálati módszerek eredményei között. Megfelelő adaptáció által a távérzékelés és adatfeldolgozás módszertani fejlesztései a gyakorlati mezőgazdaságban is alkalmazhatók lehetnek, növelve a talajtérképek pontosságát, valamint csökkentve a kapott információ költségét.

A talajtérképezés erőforrás-szükségletének csökkentésére további lehetőséget nyújt a laboratóriumi vizsgálatok újabb, idő- és költséghatékonyabb módszerekkel való részleges kiváltása. Részben a hagyományos analitikai módszerek optimalizálására, részben pedig azok kiváltására alkalmas módszerek közül számos eredményt publikáltak a spektroszkópia területén (KNOX et al., 2015; RAMESH et al., 2019; VISCARRA ROSSEL et al., 2006; WARURU et al., 2015).

A reflektancia spektroszkópia gyors, megbízható és költséghatékony vizsgálati módszer, jól alkalmazható a hagyományos módszerek kiegészítésére (GE et al., 2014). A két legtöbbet kutatott hullámhossztartomány a középhullámú (MIR: 2500–25000 nm), valamint a látható és közeli infravörös (VIS-NIR: 350–2500 nm). Mivel a két spektrum-tartomány, valamint a detektálásukhoz szükséges eszközök jelentősen eltérnek, ezért a szakirodalomban többnyire külön használatukról olvashatunk (CSORBA et al., 2012; REEVES, 2010). A VIS-NIR (Látható-Közeli Infravörös) hullámhossz-tartomány előnye, hogy az ebben a tartományban mérő műszerek költsége alacsonyabb, továbbá a hordozható eszközök elérhetősége miatt a terepi alkalmazása is elterjedt. Ezen felül a már rögzített eredményekből levont következtetések nagy területen alkalmazhatók, ugyanis ebben a hullámhossz-tartományban mérnek az optikai tartományban rögzítő műholdak is (Landsat, Sentinel). Ezzel szemben a MIR nagyobb mennyiségű információt szolgáltat a talajról, de érzékenyebb műszerezése miatt korlátozott a terepi alkalmazhatósága, valamint a műholdképek ebben a tartományban csak részlegesen szolgáltatnak információt. Jelen munka során használt MIR reflektancia elnyelődési sávjai szoros összefüggésben állnak a talaj szerves anyagának mennyiségével, így jól

alkalmazhatók a szerveszén tartalom becslésére (ATEKU, 2014; BELLON-MAUREL & MCBRATNEY, 2011).

Az alkalmazott MIR spektroszkópia által olyan típusú adathalmaz keletkezik, ami a hagyományos statisztikai módszerekkel nehezen értelmezhető, mivel a spektrumok sajátosságait lineáris módon nem lehet specifikus kémiai komponensekhez rendelni (PELTRE et al., 2014). Az eredményül kapott spektrumok kvantitatív feldolgozásához, helyes értelmezéséhez szükség van olyan matematikai és statisztikai módszerekre, amelyekkel a nem lineáris összefüggések is feltárhatók. Erre a kihívásra válaszul az utóbbi évtizedben, a begyűjtött adatok mennyiségével és összetettségével párhuzamosan fejlődött a többváltozós statisztikai adatelemzés, illetve a kemometria (HÉBERGER, 2008). Alkalmazásukkal a gyors, kémiai mintaelőkészítést nem igénylő MIR vizsgálat eredménye alapján nagy pontossággal becsülhető a hagyományos módon nem mért minták szerves széntartalma, ezáltal az adott adathalmaz vizsgálati ideje és költségei is jelentősen redukálhatók.

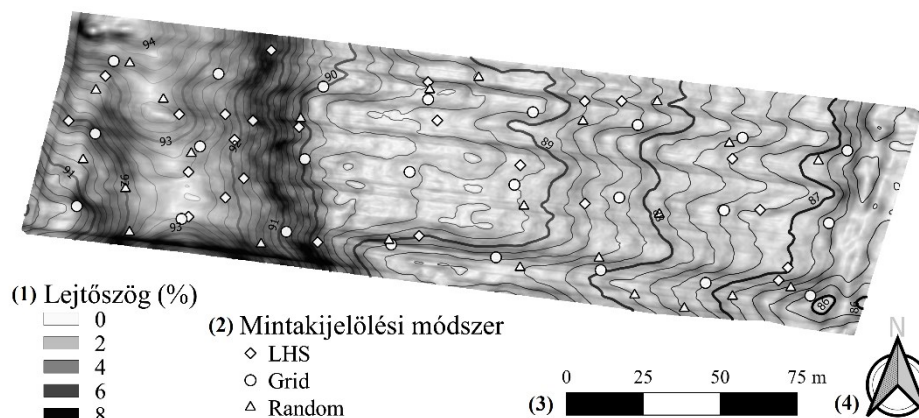
Jelen kutatás során a vizsgált terület szerveszén-felmérésének mintázását végeztük, célul tűzve ki a begyűjtött talajminták laboratóriumi vizsgálatának költségcsökkentését. Ennek érdekében a talajminták hagyományos módszerekkel történő vizsgálatát MIR spektroszkópiára alapozott kemometriai becslési modellel egészítettük ki. Vizsgáltuk a gyűjtött talajmintákból a különböző mintavételi stratégiák szerinti mintakiválasztás hatását a modell sikerességére. Vizsgáltuk továbbá a különböző modellek kalibrációjához használt minták számának hatását a modell sikerességére. A felállított modellek eredményeinek statisztikai vizsgálataiból információt kaptunk az adott mintakiválasztási módszer tulajdonságairól. Továbbá arról, hogy mely módszer alkalmas akár alacsonyabb kalibrációs mintaszámmal a talaj szerveszén viszonyainak felmérésére, ezáltal a felmérés költségeinek csökkentésére.

A vizsgálatban szereplő terület mérete és talajtani heterogenitása, valamint a mintavételi mélység jelentősen befolyásolja a mintavételi stratégia sikerességét a vizsgált terület változatosságának rögzítésére, összességében tehát a becslési modellek gyakorlati alkalmazhatóságát. Ezért választottuk azt az általánosan legkisebb egybefüggő egységet (egységesen művelt táblarész), amellyel a talajművelési gyakorlatban jellemzően találkozhatunk.

Anyag és módszer

Mintaterület bemutatása és mintavételezés

A 4,62 ha nagyságú mintaterület Érsekcsanád község közigazgatási területén, a Duna korábbi árterében található, így talajának kialakulásában jelentős szerepet játszottak a folyam egykori elöntései (MAROSI & SOMOGYI, 1990). Ennek megfelelően a terület talaja magán hordozza az öntés bélyegeket, jelentős textúraváltozások észlelhetők, vertikálisan és horizontálisan is. A heterogenitást tovább növeli a 3 m tengerszint feletti magasságcsökkenés a terület keleti és nyugati oldala között. A mélyebben elhelyezkedő részekben a közeli (1,5–2 m, évszaktól függően) talajvíz nyomán 1 m mélység alatt redukált viszonyok tapasztalhatók.



1. ábra.

Az összes mintavételi pont ábrázolva a mintaterületen

Mivel az eredmények további tanulmányok elkészítését is szolgálták, ezért olyan kompromisszumos mintavételi megoldást kerestünk, ami által a gyűjtött minták több kutatási cél megvalósítására is alkalmasak lehetnek. A mintavétel során 3 gyakori mintavételi módszer kombinációját alkalmaztunk. A három eljárás a (1) véletlenszerű (random), (2) a rács (grid), illetve (3) a domborzati adatokon alapuló LHS (Latin Hypercube Sampling – Latin-hiperkocka mintavétel) volt, 24-24 mintaponttal (1. ábra). Az utóbbi módszerhez szükséges domborzati térképet saját drónos felvétel felhasználásával készítettük. A grid és random pontok kijelölését a QGIS térinformatikai szoftverrel, a Latin Hypercube pontokét pedig az R szoftver „clhs” csomagjával végeztük el (ROUDIER, 2011). A 3 mintakijelölési módszerből összesen létrejött 72 mintavételi pont mindegyikén 100 cm³-es bolygatatlan mintát (Eijkelkamp 07.53.SC) vettünk, három mélységben: 0–20 cm, 20–50 cm és 50–100 cm. A talajminták minden esetben a mintavételi mélységtartomány közepéből származtak.

Labor analitikai és spektroszkópiai vizsgálatok

A begyűjtött talajmintákat – légszárazra történő szárításukat követően – előkészítettük további laborvizsgálatokra. A szemmel látható, nagyobb méretű növényi maradványok eltávolításra kerültek, majd a mintát 0,4 mm-es szitán hiánytalanul áttörtük. Homogenizálást követően kivett kevert részmintát 0,2 mm-es szitán törtük át, achát mozsár segítségével.

Az elválasztott részminták teljes szerveszén-tartalmát nedves oxidációs mérési eljárással (Walkley-Black) határoztuk meg, 3 ismétlésben (WALKLEY, 1947). A spektrális méréseket egy kifejezetten szilárd és porszerű anyagok vizsgálatára kifejlesztett DRIFT modulval felszerelt Bruker Alpha II típusú Fourier-transzformációs közép-infravörös spektroszkóppal (FTIR) végeztük. A mérések

során követett protokoll alapja az AfSIS (Africa Soil Information Service) projektet is koordináló Word Agroforestry Center Soil-Plant Spectral Diagnostics Lab által kidolgozott Standard Operation Procedures dokumentuma (ATEKU, 2014). A mérés során a talajminták diffúz spektrális reflektanciáját rögzítettük, a 400-4000 1/cm-es frekvencia-tartományban, mintánként 3 ismételtsben. A mérések során használt OPUS (Optics User Software) mind a 3 ismételés esetében egyenként 24 mérés átlagából származtat eredményt, az eredményül kapott spektra felbontása 2 1/cm. A háttérképeket arannyal bevonatolt referenciamintán végeztük, 3 mérésenként, így a mérés során minimalizáltuk a környezet által okozott sugárzási zaj, hőmérséklet és páratartalom-változás esetleges torzító hatását.

A Walkley-Black módszer eredményeként kapott szervesszén (TOC% - Total Organic Carbon - Összes szervesszén-tartalom) eredmények statisztikai feldolgozását a Past (<https://folk.uio.no/ohammer/past/>) szoftverkörnyezetben készítettük el. Az adatsokaságból a mintavételi mélység szerint létrehoztunk 3 csoportot. Az általános statisztikai mutatók kiszámítását követően a különböző talajmélységek értékeinek összehasonlítására Tukey-tesztet végeztünk.

Spektrális adatok elő-feldolgozása

A nyers MIR spektrum önmagában közvetlenül nem alkalmas a szerves széntartalom meghatározására. A szerves anyagoknak ugyan jól definiált elnyelési sávjai vannak, a sávok „intenzitása”, illetve „mélysége” nem kapcsolható lineárisan az egyes talajparaméterek értékeihez (BISHOP et al., 1994). Ezért – az értelmezéshez – szükséges a felvételezett spektrális adatok előfeldolgozása. Ezt követően kemometria modellt állítunk fel a minta populáció egy jól definiált részének lemerésével (kalibráció). Így becsülhetjük a kiválasztott spektrumok felhasználásával a vizsgált paramétert, a mintasor összességére vonatkozóan.

A spektrális adatok feldolgozása, illetve kiértékelése R szoftverrel (R CORE TEAM, 2017) történt. A feldolgozáshoz szükséges szkriptet a következő szoftvercsomagok felhasználásával hoztuk létre: „soil.spec” (SILA et al., 2014); „tidyverse” (WICKHAM, 2017); „pls” (MEVIK et al., 2019); „prospectr” (STEVENS & RAMIREZ-LOPEZ, 2013); „resemble” (RAMIREZ-LOPEZ & STEVENS, 2016); „caret” (KUHN et al., 2019).

A spektroszkópiai mérések eredményeinek előfeldolgozása során elsőként az ismételésekből származó spektrumok átlagolását és a spektrumokon mutatkozó „fűrészfog” jelenség csökkentését végeztük el. Utóbbit egy a spektrumokon elvégzett, mozgó átlag számításra alapozott simító eljárással végeztük el.

A kemometria modellezés megkezdése előtt a végső lépés az esetleges mérési hiba miatt, vagy a minta más populációba tartozása miatt előforduló kiugró értékek kiszűrése. Ehhez a spektrális adatbázis alapján, minden mintára kiszámított főkomponens faktorértékeken elvégzett távolságszámítási módszerek eredménye alapján meghatároztuk a minták egymáshoz való hasonlóságának/egymástól való különbözőségének mértékét. Távolságszámítási módszerként a minták egymástól való távolságát a Mahalanobis és H távolság értékek megállapításával határoztuk

meg. Az eredményként kapott diagramok vizuális értékelése alapján megállapított határértékek definiálásával szűrtük ki a kiugró értékű mintákat.

Kalibrációs mintakiválasztás és kemometriai modellezés

Az összes minta MIR spektrumának elő-feldolgozását követően azok közül az egész sokaságra reprezentatív mintákat választottuk ki, a későbbi modell felállításához. A többváltozós adatok elemzése során számos algoritmus létezik a kalibrációs minták kiválasztására, a leggyakrabban alkalmazott módszerek a KSS (Kennard-Stone sampling – Kennard-Stone mintavétel), KMS (K-means sampling – K-közép mintavétel) és LHS (Latin Hypercube sampling – Latin hiperkocka mintavétel) (RAMIREZ-LOPEZ et al., 2015). A KSS az infravörös spektroszkópia területén a legtöbbet használt módszer, azokat a mintákat választja ki a sokaságból, amelyek a változók alapján a legnagyobb távolságra vannak egymástól. A KMS ezzel szemben azonos tulajdonságaik szerint csoportokba rendezi az adatpontokat, ezt követően véletlenszerűen választ mintákat az egyes csoportokból. Az LHS a bevitt változókat egy normalizált latin négyzetbe rendezi, amiből úgy jelöl ki pontokat, hogy a választott minták legjobban fejezzék ki a többváltozós rendszer belső eloszlását (NG et al., 2018). Ebben a lépésben megtörtént tehát a referencia minták kiválasztása a KSS, KMS és LHS módszerek mindegyikével. Az ily módon definiált mintákat használtuk a kemometriai modell kalibrálására, így vizsgálva a talajparaméter (szerves szén) és a spektrum közötti kvantitatív összefüggést. A létrehozott modellt ezt követően elsődlegesen validálni szükséges további, ismert szervesanyag-tartalmú, a kalibrációtól független minták bevonásával. Mivel az eredmények összefüggései specifikusak az adott mintaterületre, azok értékelése során a levont következtetések is mindig a vizsgált terület sajátosságait tükrözik. Ily módon tehát, jelen kézirat eredményei elsősorban a vizsgálatban szereplő talajviszonyok mellett értelmezhetők, más, jelentősen eltérő talajtípusok esetében a feltárt összefüggéseket mindenképp szükséges felülvizsgálni.

Általános esetben a mintakijelölési eljárások az adott adatsokaság leginkább reprezentatív mintáinak kiválasztására alkalmasak. Jelen vizsgálat során azonban a spektrumokon felül rendelkezésre állt minden mintára vonatkozóan a szerves széntartalom érték is. Eszerint az említett 3 mintakiválasztási módszer részletesebb vizsgálatához az összes talajminta TOC értékének felhasználásával végeztünk kalibrációs/validációs lépéseket, változó arányban. Mindegyik módszer esetében 9 lépéses vizsgálatot végeztünk. Ennek során a teljes mintapopuláció 10%-tól kiindulva minden lépésben 10%-kal növeltük a kalibrációs minták mennyiségét, egészen 90%-ig. Így tesztelhetjük a KSS, KMS és LHS módszerek által kiválasztott minták felhasználásával létrehozott modellek teljesítményét. A tesztelés során várhatóan az alacsony mintaszámmal (<30%) történő kalibrálás és validálás egyaránt a modell bizonytalanságát okozhatja. Ezt a jelenséget a modellek teljesítményvizsgálatának elemzése során szükséges figyelembe venni.

A kemometriai modellt a PLSR (Partial Least Squares Regression – Parciális legkisebb négyzetek regresszió) módszer alkalmazásával állítottuk fel, mely a gyakorlatban az egyik leggyakrabban alkalmazott modellezési eljárás (DUCKWORTH,

1998). A számos kemometriai módszer közül a PLSR alkalmazható a leghatékonyabban prediktív modellek létrehozására, amikor a független (X) változók (a spektrum hullámhosszain mutatkozó spektrális értékek) száma meghaladja a minták számát, és változók között jelentős multikollinearitás mutatkozik, mely a spektrális adatokra fokozottan igaz. A módszer szoros rokonságban áll a főkomponens regresszióval (Principal Component Regression – PCR), azonban míg a PCR a spektrális adatokon elvégzett főkomponens analízis során számított faktorértékeket használja csak független változókként, addig a PLSR törekszik a spektrális értékek és a talajparaméterek közötti kovariancia maximalizálására. Ezt a célt a PLSR algoritmus egymást követő faktorok kiválasztásával, az X és Y változók mátrixának faktorértékekké és faktorsúlyokká való szétbontásával éri el úgy, hogy az első néhány faktor kifejezze az X és Y változókban jelenlévő változékonyság legnagyobb hányadát. Ismeretlen összetételű minták tulajdonságainak becslése a minta spektrumának PLSR modellbe vonásával, a mintára kiszámított faktorsúly és faktorérték kombinációjával érhető el.

A többváltozós eredmények statisztikai vizsgálatán túl, a kémiai analitikai eredmények értelmezésére alkalmazható a felállított kemometriai modell. A kiválasztott kalibrációs minták felhasználásával egy olyan matematikai-statisztikai alapokon nyugvó modellt hoztunk létre, mely használatával a kiválasztott minták szervesszén adatai alapján becsülhettük a teljes adatsokaság spektrális alapú szerves széntartalmát.

Kemometriai modellek validációja

Vizsgáltuk a különböző mintaszámmal elvégzett kalibráció hatását a modellek statisztikai mutatóira. Ehhez az egyes mintakijelölési módszerekkel kiválasztott mintákból létrehozott modellek statisztikai paramétereinek változását (RMSE – Root Mean Squared Error – Átlagos négyzetes gyökeltérés; R^2 – Determinációs együttható) vizsgáltuk, a kalibrációs minták számának változtatása során. A modell teljesítményének jellemzésére kifejezetten jelen kutatás számára kidolgoztunk egy olyan összegzett statisztikai mutatót, amelynek eredménye egy változóban fejezi ki az R^2 és az RMSE értékeit, normalizált skálán. A létrehozott változó képlete: $(R^2 - RMSE) / (R^2 + RMSE)$. Ezen statisztikai mutató létrehozásához használt egyenlet alapja a távérzékelési gyakorlatban gyakran alkalmazott NDVI index (Normalizált vegetációs index - Normalized Difference Vegetation Index) készítése során használt képlet (PETTORELLI, 2013). Az R^2 növekedése, illetve az RMSE csökkenése növeli a mutató értékét, így az összegzett statisztikai mutató nagyobb értéke esetén a tesztelt modell pontossága is nagyobb, tehát alacsonyabb hibával alkalmas az adatsokaság kifejezésére.

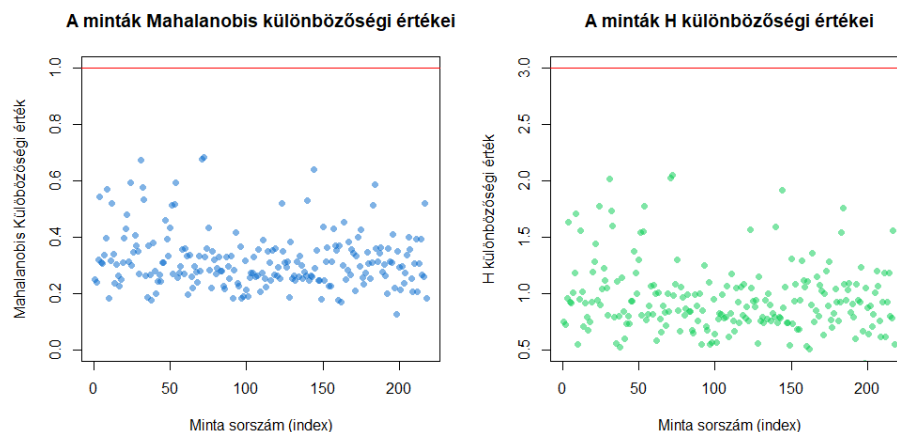
Eredmények

1. táblázat

A mért szervesszén értékek statisztikai mutatói, illetve összehasonlítása

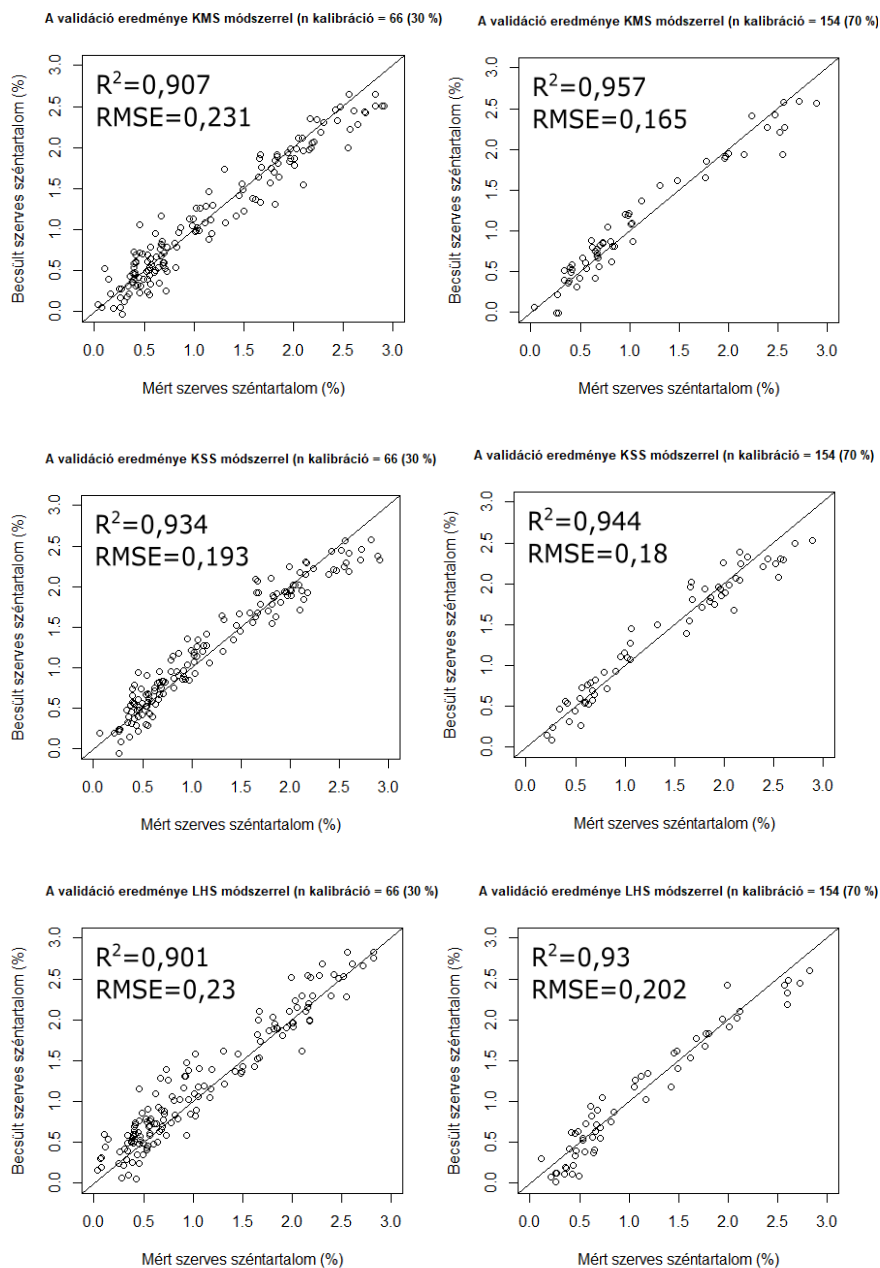
Mélység (1)	0–20 cm	20–50 cm	50–100 cm	Tukey's Q értékek a diagonál alatt, a p értékek felette találhatók. (2)			
Min/max (%)	0,38 / 2,91	0,19 / 2,82	0,03 / 1,32				
Átlag (3)	1,46	1,29	0,61				
Standard hiba (4)	0,10	0,09	0,04		0–20 cm	20–50 cm	50–100 cm
Szórás / Szórás-négyzet (5)	0,86 / 0,74	0,77 / 0,6	0,31 / 0,1	0–20 cm		0.2753	1.115E-11
Medián	1,64	1,25	0,67	20–50 cm	2.175		6.098E-08
CV %	58,66	60,09	50,81	50–100 cm	10.42	8.245	

A Walkley-Black teszt eredményeként kapott szervesszén (TOC %) értékek statisztikai vizsgálati eredménye jól igazolja az adathalmaz heterogenitását. Az 1. táblázatban szerepelnek a 3 mélység szerint különválasztott adatszoportok értékeinek alapvető statisztikai mutatói. Ez alapján elmondható, hogy a vizsgált 216 minta jelentős szórást és varianciát mutat. Bár a mintákat a további vizsgálatok során nem választottuk külön mélység szerint, a mintasokaság így módon történő statisztikai vizsgálata információt szolgáltat az adatok belső elrendezetségéről és varianciájáról, amely tulajdonságok a felhasznált modellek viselkedését alapvetően befolyásolják.



2. ábra

A mért spektrumok különbség értékei.



3. ábra

Validációk eredménye a különböző módszereknél (30% és 70% kalibrációs minta).

A kiugró értékű minták kvantitatív alapon történő kiszűrését a Mahalanobis és H távolságok számításával végeztük, amelyhez tapasztalati úton, valamint szakirodalmi példa alapján meghatározott küszöbértéket használtunk. A Mahalanobis távolságok esetében az 1-nél, a H távolságok esetén a 3-nál nagyobb különbözőségi értékkel rendelkező mintákat tekintettük kiugró értéknek (MATAMALA, 2019). A diagramok alapján tehát nincs kiugró értékű minta az általunk felvételezett spektrális adatokban (2. ábra).

A részletes elemzés során a KSS, KMS és LHS mintavételezési módszerek becslési hatékonyságát teszteltük a kalibrációs mintaszám függvényében. A validálás során a modell megbízhatóságát egy független, a kalibrációból kimaradó mintasor alkalmazásával ellenőrizzük. A modellek részletesebb feltárása érdekében 9 lépéses kalibrációs / validációs vizsgálatot végeztünk, melynek során 10%-tól 90%-ig 10%-os lépésekben növeltük a kalibrációs minták arányát, a validációs mintákkal szemben. A PLSR kemometriai elemzések eredményeként létrejövő 27 szóródási diagram a modell kalibráció és validáció eredményeit mutatja be. A 30%-nál kevesebb minta felhasználásával történő kalibrálás, illetve validálás bizonytalansággal terheli a modell felállítását/tesztelését, ezért ezen szélső értékek mentén (az első esetben a kalibrációban felhasznált minták száma 30%, második esetben 70%.) választottuk ki az ábrázolt szóródási diagramokat (3. ábra).

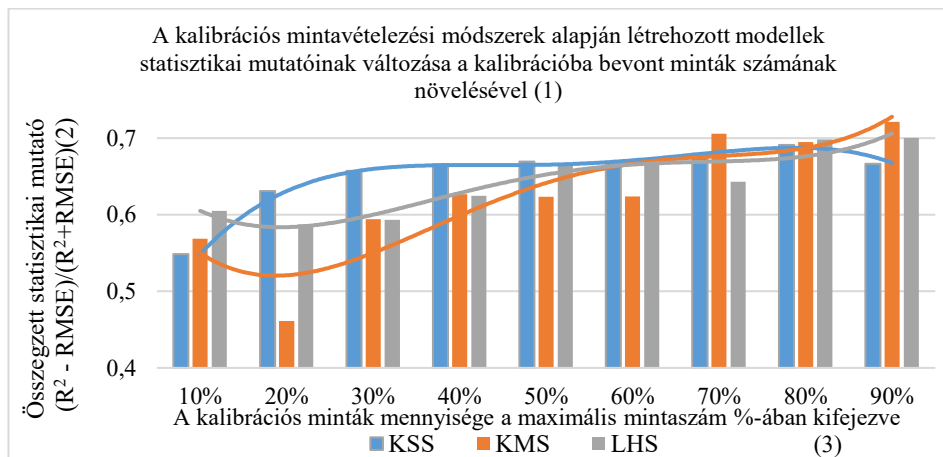
A különböző arányú kalibrációkat követő validációs lépésben a modell megbízhatóságát annak statisztikai mutatói (R^2 és RMSE) definiálják (2. táblázat). Az R^2 érték azt fejezi ki, hogy a becsült változók (talajparaméterek) változatossága milyen mértékben mutatkozik a becslések alapját képező (spektrális) adatokban. Az RMSE értéke pedig megadja, hogy a modell a validációs lépése során mekkora közepes négyzetes eltéréssel becsülte az adott talajparamétert. Az RMSE dimenziója minden esetben megegyezik az adott talajparaméter mértékegységével, alacsony értéke a becslések kisebb hibáját jelenti.

2. táblázat

A becslési modellek megbízhatósági vizsgálatának eredménye.

	A kalibrációban részt vevő minták száma a teljes mintaszám %-ában kifejezve (1)								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
KSS R^2	0,885	0,924	0,934	0,938	0,939	0,936	0,944	0,952	0,93
KSS RMSE	0,258	0,209	0,193	0,188	0,186	0,187	0,18	0,174	0,186
KMS R^2	0,895	0,835	0,907	0,92	0,927	0,906	0,957	0,944	0,969
KMS RMSE	0,246	0,308	0,231	0,211	0,215	0,21	0,165	0,17	0,157
LHS R^2	0,914	0,904	0,901	0,922	0,937	0,939	0,93	0,951	0,948
LHS RMSE	0,225	0,235	0,23	0,213	0,187	0,187	0,202	0,169	0,167

Az abszolút tekintetben vett legalacsonyabb és legmagasabb értékű összegzett statisztikai mutatók is a KMS módszerhez tartoznak, értékük 0,46 és 0,72. Az ábrán látható, hogy a modellek pontossága már 10%-os kalibráció esetén is elfogadható, azonban a kalibrációs mintaszám 30%-ig való növelésével jelentős ingadozást észleltünk (4. ábra). A kalibráció 30 és 70% közötti részén a becslés pontosságának általános növekedése látható, majd 80 és 90% esetén újra ingadozás tapasztalható.



4. ábra

A modellek becslési pontosságának összegzett statisztikai mutatói.

Az egyes modellekhez tartozó görbék telítődése eltérő kalibrációs mintaszámnál következik be. Ez az érték a KSS esetén 30%, a KMS esetében 60%, az LHS esetében pedig 50%. Az eredményeket összegezve értelmezhetjük a felhasznált modellek általános viselkedését. A KSS alacsony mintaszámú kalibráció esetén alacsony reprezentativitással rendelkezik, azonban a kalibrációs mintaszám növelésével ugrásszerűen nő a becslés pontossága. A KMS és LHS módszerekre együttesen elmondható, hogy 10%-os kalibráció esetén nagyobb reprezentativitással jellemezhetők, azonban a 20 és 30%-nál tapasztalható prediktív bizonytalanságot követően ugyanezt a reprezentativitást csak 40%-os kalibrációs mintaszám esetén haladják meg. A 10%-os kalibráció melletti magas reprezentativitás nagy valószínűséggel nem a modell sikerességének tulajdonítható.

Következtetések

A szervesszen adatokban jelentkező magas szórás értékek és az adatokon belüli csoportok (mélység) igazolt különbsége kapcsán kijelenthető, hogy a modellek teljesítményét az ehhez szükséges heterogenitás mellett vizsgáltuk. A gyakorlatban hasonló méretű tábláról begyűjtött mintahalmazon belül jellemzően hasonló mértékű, vagy alacsonyabb heterogenitás értékek fordulnak elő. Mivel a tesztelt módszerek

hatékonysága függ a heterogenitástól, ezért az adatpontok szórásának és varianciájának növelésével az elkészített modell hatékonysága várhatóan csökken.

A mintakiválasztási módszerek statisztikai értékelése során az alacsony mintaszámmal (<30%) végzett kalibráció esetén tapasztalt ingadozás a modellek bizonytalanságára utal. Ez alól kivétel a KSS módszer, melynek hatékonysága 10%-tól 80%-ig folyamatosan nőtt. A magas mintaszámú (>70%) kalibráció esetében tapasztalt fluktuációt a validációs minták között előforduló szélső értékek okozzák. Ez – valamint a pontosságot jelző mutatók 30 és 70% közötti folyamatos javulása – alapján kijelenthető, hogy a vizsgált mintakijelölési módszerek ez utóbbi mintaszám-tartományon belül alkalmazhatók a legnagyobb hatások mellett.

Tapasztalható volt a mintakiválasztási módszerek tesztelése során a becslési pontosságot bemutató görbék telítődése. A telítődési pontok azt a legkisebb kalibrációs mintaszámot jelzik, amely értéktől a modell reprezentativitásának görbéje ellaposodik. Ezek alapján kiválaszthatjuk az alacsony kalibrációs mintaszám mellett leghatékonyabb, valamint az abszolút értelemben vett leghatékonyabb módszereket.

A legkisebb költség mellett (legkisebb kalibrációs mintaszám) reprezentatív mintakijelölési módszer a KSS volt, 30% kalibrációs mintaszámmal. A legnagyobb reprezentativitást a KMS alkalmazásával, a minták 90%-val való kalibrációjával érték el.

Fontos megjegyezni, hogy a kapott eredmények az adott vizsgálati terület talajára specifikusak. A kapott eredmények jól használhatók tábla szintű talajtérképezés esetén, azonban jelentősen eltérő talajadottságú, vagy nagyobb léptékű, heterogénebb területekhez további vizsgálatokra van szükség. A kapott eredmények nagyobb léptékű térképezés során való felhasználásához szükséges a vizsgált módszerek nagyobb területen, nagyobb heterogenitás melletti tesztelése.

Összefoglalás

Szervesszén térképezést segítő módszertani kutatásként vizsgáltuk egy szántóföldi művelés alatt álló terület, 3 mélységből származó mintáinak MIR reflektanciáját, illetve szervesszén tartalmát (Walkley-Black). Ezt követően a spektroszkópia mérések eredményeit használtuk a talaj szervesszén-mennyiségének (TOC %) becslésére. Tettük ezt 3 mintakijelölési módszer (Kennard-Stone Sampling - KSS, K-means Sampling - KMS, Latin Hypercube Sampling - LHS) bevonásával, az így kijelölt kalibrációs mintákkal a PLSR modell segítségével becslést végeztünk az adathalmaz további értékeire. Annak érdekében, hogy tábla szintű szervesszén meghatározás során teszteljük becslési pontosságukat, a modellek reprezentativitását – különböző validációs/kalibrációs arány esetén – statisztikai mutatókkal (R^2 , RMSE) ellenőriztük.

Az eredményekben részleteiben vizsgáltuk a különböző becslési modellek reakcióját eltérő arányú kalibráció és validáció esetén. A modellek R^2 és RMSE értékei alapján kijelöltük, hogy mely modellek működtek pontosan még alacsony kalibráció esetén is, illetve abszolút értelemben véve melyik modell volt leghatékonyabb.

Az összehasonlítás eredményeként kijelenthető, hogy az általunk vizsgált talajkörülmények között a 30% alatti, valamint a 70% feletti mintaszámú kalibráció a mintakijelölési módszerek megbízhatóságának ingadozását eredményezte. Az összes minta 30%-val történő kalibráció esetén legjobb eredményt a KSS adta, így ez tekinthető a leggazdaságosabb módszernek. Az abszolút értékben vett legkisebb hibát a K-means sampling eredményezte, a minták 90%-val történő kalibrációt követően.

Kijelenthető, hogy az alkalmazott módszertan esetünkben alkalmas volt – a reprezentativitás megtartása mellett – a szükséges minták számának, ergo a táblaszintű szervesszén-felmérés költségeinek csökkentésére. Továbbá a mintakijelölési módszerek becslési hatékonyságának összehasonlítására is megfelelt az általunk alkalmazott statisztikai vizsgálat. A módszertan a jövőben kiinduló alapja lehet hasonló jellegű kutatásoknak, valamint tábla szintű szervesszéntérképek elkészítésének. A szélesebb körű alkalmazást megelőzően a modelleket nagyobb varianciájú adathalmazok esetén is tesztelni szükséges.

Köszönetnyilvánítás

A jelenlegi publikációban közölt eredményekhez kapcsolódó kutatómunka az EFOP-3.6.1-16-2016-0016 pályázati projekt keretein belül valósult meg. A kutatást továbbá az Innovációs és Technológiai Minisztérium támogatta a Tématérületi Kiválósági Program 2020, Intézményi Kiválósági Alprogram (TKP2020-IKA-12) keretében a Szent István Egyetem vízzel kapcsolatos kutatásainak körében.

Irodalom

- ATEKU, D. (2014). Standard Operating Procedure - Mid-Infrared Diffuse Reflectance Spectroscopy Method for Analysing Soils using Bruker Alpha Spectrometer. World Agroforestry Center, Code: METH07V01. 10 pp.
- BALLABIO, C., PANAGOS, P., MONATANARELLA, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*. **261**. 110–123. <https://doi.org/10.1016/j.geoderma.2015.07.006>
- BELLON-MAUREL, V., MCBRATNEY, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. *Soil Biology and Biochemistry*. **43**. (7) 1398–1410.
- BISHOP, J. L., LANE, M. D., DYAR, M. D., BROWN, A. J. (1994). Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite–serpentines, chlorites and micas. *Clay Minerals*. **43**. 53–54.
- CHENU, C., ANGERS, D., BARRÉ, P., DERRIEN, D., ARROUAYS, D., BALESSENT, J. (2019). Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. *Soil and Tillage Research*. **188**. 44–52.
- CSORBA Á., LÁNG V., FENYVESI L., MICHÉLI E. (2012). Reflektancia spektroszkópia alkalmazása talajok szervesszén- és CaCO₃-tartalmának becslésében. *Agrokémia és Talajtan*. **61**. (2) 277–290.

- DEERY, D., JIMENEZ-BERNI, J., JONES, H., SIRAUT, X., FURBANK, R. (2014). Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy*. <https://doi.org/10.3390/agronomy4030349>
- DOBOS E., MICHÉLI E., MONTANARELLA L. (2007). The population of a 500-m resolution soil organic matter spatial information system for Hungary. *Developments in Soil Science*. **31**. 487–495.
- DOBOS E., SERES A., BOCK M., KÖTHE, R., DAROUSSIN J., VAN ENGELEN V. (2010). Landform mapping for SOTER at scale 1:1 million using SRTM-DEM. 19th World Congress of Soil Science, Soil Solutions for a Changing World. 1–6 August 2010, Brisbane, Australia. DVD published, pp. 68–71.
- DOBOS E., VADNAI P., BERTÓTI R. D., KOVÁCS K., MICHÉLI E., SZEGI T., FULLAJTÁR E., PENIZEK V., SWITONIAK M. (2014). Új WRB alapú validációs adatbázis és validációs módszertan Közép-Európára, ValiDat.DSM. *Agrokémia és Talajtan*. **63**. (2) 393–408.
- DOBOS E., VADNAI P., KOVÁCS K., LÁNG V., FUCHS M., MICHÉLI E. (2019). A novel approach for mapping WRB soil units – A methodology for a global SOTER coverage. *Hungarian Geographical Bulletin*. **68**. (2) 157–175.
- DUCKWORTH, J. H. (1998). Spectroscopic quantitative analysis. In: *Applied Spectroscopy* (Eds.: Workman, J. & Springsteen, A.) Academic Press. San Diego, California. 93–163.
- GE, Y., THOMASSON, J., MORGAN, C. (2014). Mid-infrared attenuated total reflectance spectroscopy for soil carbon and particle size determination. *Geoderma*, **213**. 57–63.
- HÉBERGER, K. (2008). Chapter 7 - Chemoinformatics—multivariate mathematical-statistical methods for data evaluation. *Medical Applications of Mass Spectrometry*. 141–169.
- KNOX, N., GRUNWALD, S., MCDOWELL, M., BRULAND, G., MYERS, D., HARRIS, W. (2015). Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma*. **239–240**. 229–239.
- KUHN, M., WING, J., WESTON, S., WILLIAMS, A., KEEFER, C., ENGELHARDT, A., COOPER, T., MAYER, Z., KENKEL, B., R CORE TEAM. (2019). caret: Classification and Regression Training. R package version 6.0-84. <https://CRAN.R-project.org/package=caret>
- LAL R. (2010). Enhancing Eco-efficiency in Agro-ecosystems through Soil Carbon Sequestration. *Crop Science*. **50**. 120–130.
- LAL R. (2013). Soil carbon management and climate change. *Carbon Management*. **4**. (4). 439–462.
- LAL R. (2017). Soil Carbon Impacts on Functionality and Environmental Sustainability. In: Erşahin S., Kapur S., Akça E., Namlı A., Erdoğan H. (Szerk.) *Carbon Management, Technologies, and Trends in Mediterranean Ecosystems. The Anthropocene: Politik—Economics—Society—Science*. vol. 15. Springer, Cham
- LAL R., NEGASSA W., LORENZ K. (2015). Carbon sequestration in soil. *Environmental Sustainability*. **15**. 79–86.

- MAROSI S., SOMOGYI S. (szerk., 1990). Magyarország kistájainak katasztere I-II. MTA Földrajztudományi Kutató Intézet, Budapest. 1023 p.
- MATAMALA, R., JASTROW, J. D., CALDERÓN, F. J., LIANG, C., FAN, Z., MICHAELSON, G. J., PING, C. L. (2019). Predicting the decomposability of arctic tundra soil organic matter with mid infrared spectroscopy. *Soil Biology and Biochemistry*. **129**. 1–12.
- MEVIK, B-H., WEHRENS, R., LILAND, K. H. (2019). pls: Partial Least Squares and Principal Component Regression. R package version 2.7-1. <https://CRAN.R-project.org/package=pls>
- NG W., MINASNY B., MALONE B., FILIPPI P. (2018). In search of an optimum sampling algorithm for prediction of soil properties from infrared spectra. *PeerJ*. 6:e5722; DOI 10.7717/peerj.5722.
- PAHLAVAN RAD, M. R., TOOMANIAN, N., KHORMALI, F., BRUNGARD, C. W., KOMAKI, C. B., BOGAERT, P. (2014). Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2014.04.036>
- PÁSZTOR L., LABORCZI A., SZATMÁRI G., TAKÁCS K., BAKACSI Z., SZABÓ J., DOBOS E. (2014). Introduction of digital soil mapping techniques for the nationwide regionalization of soil condition in Hungary; the first results of the DOSoReMI.hu (Digital, Optimized, Soil Related Maps and Information in Hungary) project. In: EGU General Assembly 2014, Vienna, Austria, id.7310.
- PÁSZTOR L., LABORCZI A., TAKÁCS K., ILLÉS G., SZABÓ J., SZATMÁRI G. (2020). Progress in the elaboration of GSM conform DSM products and their functional utilization in Hungary. *Geoderma Regional*. <https://doi.org/10.1016/j.geodrs.2020.e00269>
- PÁSZTOR L., LABORCZI A., TAKÁCS K., SZATMÁRI G., DOBOS E., ILLÉS G., BAKACSI Z., SZABÓ J. (2015). Compilation of novel and renewed, goal oriented digital soil maps using geostatistical and data mining tools. *Hungarian Geographical Bulletin*. **64**. (1) 49–64.
- PÁSZTOR L., SZABÓ J., BAKACSI Zs. (2010). Application of the Digital Kreybig Soil Information System for the delineation of naturally handicapped areas in Hungary. *Agrokémia és Talajtan*. **59**. 47–56.
- PÁSZTOR L., SZABÓ J., BAKACSI Z., LABORCZI A. (2013). Elaboration and applications of spatial soil information systems and digital soil mapping at Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences. *Geocarto International*. **28**. (1) 13–27.
- PELTRE, C., BRUUN, S., DU, C., THOMSEN, I., JENSEN, L. (2014). Assessing soil constituents and labile soil organic carbon by mid-infrared photoacoustic spectroscopy. *Soil Biology and Biochemistry*. **77**. 41–50.
- PETTORELLI N. (2013). NDVI from A to Z. In: *The Normalized Difference Vegetation Index*. Szerk: PETTORELLI N., Oxford University Press. ISBN: 978-0-19-969316-0. 30-43.
- R CORE TEAM (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

- RAMESH, T., BOLAN, N., KIRKHAM, M., WIJESEKARA, H., KANCHIKERIMATH, M., SRINIVASA RAO, C., SANDEEP, S., RINKLEBE, J., OK, Y., CHOUDHURY, B., WANG, H., TANG, C., WANG, X., SONG, Z., FREEMAN II, O. (2019). Chapter One - Soil organic carbon dynamics: Impact of land use changes and management practices: A review. *Advances in Agronomy*. **156**. 1-107.
- RAMIREZ-LOPEZ, L., SCHMIDT, K., BEHRENS, T., WESEMAEL, BAS VAN, DEMATTÉ, J. A. M., SCHOLTEN, T. (2014). Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma* (2014). <http://dx.doi.org/10.1016/j.geoderma.2014.02.002>
- RAMIREZ-LOPEZ, L., SCHMIDT, K., BEHRENS, T., VAN WESEMAEL, B., DEMATTE, J. A. M., SCHOLTEN, T. (2015). Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma*. **226–227**. 140–150.
- RAMIREZ-LOPEZ, L., STEVENS, A. (2016). resemble: Regression and similarity evaluation for memory-based learning in spectral chemometrics R package version 1.2.2
- REEVES, J. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*. **158**. (1–2) 3–14.
- ROUDIER, P. (2011). clhs: a R package for conditioned Latin hypercube sampling.
- SILA, A., HENGL, T., THOMAS TERHOEVEN-URSELMANS, T. (2014). soil.spec: Soil Spectroscopy Tools and Reference Models. R package version 2.1.4. <https://www.CRAN.R-Project.org/package=soil.spec>
- SZABÓ J., PÁSZTOR L., BAKACSI Z. (2005). Egy országos, térbeli talajinformációs rendszer kiépítésének igénye, lehetősége és lépései. *Agrokémia és Talajtan*. **54**. 41–58.
- SZATMÁRI G., BARTA K., PÁSZTOR L. (2015). An application of a spatial simulated annealing sampling optimization algorithm to support digital soil mapping. *Hungarian Geographical Bulletin*. **64**. (1) 35–48.
- SZATMÁRI G., PÁSZTOR L. (2016). Geostatistika a talajtérképezésben - Szemle. *Agrokémia és Talajtan*. **65**. (1) 95–114.
- SZATMÁRI G., PÁSZTOR L. (2019). Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. *Geoderma*. **337**. 129-134, <https://doi.org/10.1016/j.geoderma.2018.09.008>.
- SZATMÁRI G., PIRKÓ B., KOÓS S., LABORCZI A., BAKACSI Z., SZABÓ J., PÁSZTOR L. (2019). Spatio-temporal assessment of topsoil organic carbon stock change in Hungary. *Soil & Tillage Research*. **195**. <https://doi.org/10.1016/j.still.2019.104410>.
- TAN K. H. (1982). Colloidal Chemistry of Organic Soil Constituents. In: *Principles of Soil Chemistry*. 4th Edition. 75–130.
- VISCARRA ROSSEL, R., ADAMCHUK, V. I. (2013). Proximal soil sensing. In *Precision Agriculture for Sustainability and Environmental Protection*. <https://doi.org/10.4324/9780203128329>

- VISCARRA ROSSEL, R., WALVOORT, D., MCBRATNEY, A., JANIK, L., SKJEMSTAD, J. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*. **131**. (1–2) 59–75.
- WALKLEY, A. (1947). A critical examination of a rapid method for determining organic carbon in soils: Effect of variations in digestion conditions and of inorganic soil constituents. *Soil Science*. **63**. 251–263.
- WARURU, B., SHEPHERD, K., NDEGWA, G., SILA, A., KAMONI, P. (2015). Application of mid-infrared spectroscopy for rapid characterization of key soil properties for engineering land use. *Soils and Foundations*. **55**. (5) 1181–1195.
- WEIL R. R., BRADY N. (2017). Soil Organic Matter. In: *The Nature and Properties of Soils*. 15th edition. Fox D., Gilfillan A. (Szerk.) pp 526–581.
- WICKHAM, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>

Representativity analysis of middle-infrared spectroscopy-based Organic Carbon assessment on field-scale by chemometric methods

*József Attila TÓTH, Réka DÖBRÖNTEY, Tamás SZEGI, Erika MICHÉLI, Ádám CSORBA

Hungarian University of Agricultural and Life Sciences, Gödöllő

Summary

Within the framework of the present research, we mapped the organic carbon content of an arable area, during which we measured the MIR reflectance and organic carbon content (Walkley-Black) of the soil samples collected from the area at three different depths. Subsequently, the results of spectroscopic measurements were used to improve the estimation of the soil organic carbon content (TOC %). Three sample selection models were involved (KSS, KMS, LHS), and with the selected calibration samples, we estimated the additional values of the data set using the PLSR model. In order to test the accuracy of estimation for a table-level organic carbon determination, the representativeness level of the models was checked with statistical indicators (R², RMSE) at different validation / calibration ratios.

In the results, we thoroughly examined the response of different estimation models with different ratios of calibration and validation. Based on the R² and RMSE values of the models, we determined which models worked precisely even at low calibration, and in absolute terms, which model was the most efficient.

As a result of the comparison, it can be stated that under the soil conditions we examined, calibration with a sample number below 30% and above 70% caused significant fluctuations in the reliability of the sampling methods. Kennard-Stone sampling (KSS) gave the most precise results for calibration with 30% of all samples, thus it is considered the most economical method. The smallest error overall was given by K-means sampling after calibration of 90% of the samples.

It can be stated that the methodology used in this study was suitable to reduce the samples required for analysis - while maintaining representativeness - therefore reducing the costs of the field-level organic carbon survey.

Furthermore, the statistical analysis we used to compare the estimation efficiency of the sampling methods was also appropriate. The methodology we use may be the basis for similar research in the future, as well as for the production of table-level organic carbon maps. Prior to wider application, models also need to be tested for higher variance datasets.

Keywords: soil carbon stock, carbon sequestration, soil survey

Tables and figures

Table 1. Statistical indicators and comparison of the measured organic carbon values. Note: (1) Sampling depth, (2) Tukey's Honestly Significant Difference under the diagonal, adherent p values above the diagonal, (3) Mean, (4) Standard Error, (5) Standard Deviation / Variance.

Table 2. Result of reliability testing of estimation models. Note: (1) Number of calibration samples (% of total number of samples)

Figure 1. Sampling points are plotted in the sample area. Legend: (1) Slope angle, (2) Sampling method, (3) scale, (4) orientation

Figure 2. The calculated distance values of the spectra. (left: Mahalanobis Distance X axis: Index number of sample; Y axis: Mahalanobis Distance value; right: Hellinger Distance X axis: Index number of sample; Y axis: Hellinger Distance value)

Figure 3. Validation results for LHS methods (30% and 70% calibration samples). (30% calibration samples – left; 70% calibration samples – right; upper row: KMS sampling, middle row: KSS sampling, lower row: LHS sampling; X axis: Measured TOC [%]; Y axis: Estimated TOC [%])

Figure 4. Cumulated statistical indicators of the estimation accuracy of the models. Note: (1) Change in cumulated statistical indicators of prediction models in relation to increasing number of samples included in the calibration (2) Cumulated statistical indicator $(R^2 - RMSE)/(R^2 + RMSE)$, (3) The number of samples used for calibration, as a percentage of all samples;

Open Access nyilatkozat: A cikk a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0>) feltételei szerint publikált Open Access közlemény, melynek szellemében a cikk bármilyen médiumban szabadon felhasználható, megosztható és újraközölhető, feltéve, hogy az eredeti szerző és a közlés helye, illetve a CC License linkje és az esetlegesen végrehajtott módosítások feltüntetésre kerülnek. (SID_1)
